

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



Trần Minh Chiến

**MÔ HÌNH TÌM KIẾM ĐOẠN NỘI DUNG
TƯƠNG ĐỒNG GIỮA CÁC VĂN BẢN KHÔNG
CÙNG NGÔN NGỮ**

KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY
Ngành: Công nghệ thông tin

HÀ NỘI - 2019

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

Nguyễn Tuấn Anh

**PHÁT TRIỂN PHẦN MỀM NHẬN DẠNG HOA
TRÊN NỀN TẢNG THIẾT BỊ DI ĐỘNG**

KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY
Ngành: Công nghệ thông tin

Cán bộ hướng dẫn: TS. Nguyễn Thị Ngọc Diệp

Cán bộ đồng hướng dẫn: PGS.TS. Nguyễn Việt Hà

HÀ NỘI - 2019

**VIETNAM NATIONAL UNIVERSITY, HA NOI
UNIVERSITY OF ENGINEERING AND TECHNOLOGY**

Nguyen Tuan Anh

**DEVELOPING A FLOWER RECOGNITION
SOFTWARE ON MOBILE DEVICE PLATFORM**

**BACHELOR'S THESIS
Major: Information Technology**

Supervisor: Dr. Nguyen Thi Ngoc Diep

Co-Supervisor: Assoc. Prof. Nguyen Viet Ha

HANOI - 2019

LỜI CẢM ƠN

Lời đầu tiên cho tôi xin được gửi lời cảm ơn chân thành và sâu sắc nhất tới TS. Nguyễn Thị Ngọc Diệp, PGS. TS. Nguyễn Việt Hà, ThS. Nguyễn Ngọc Khương những người đã hướng dẫn và chỉ bảo tận tình nhất cho tôi trong suốt quá trình hoàn thành khóa luận tốt nghiệp này.

Tôi xin được gửi lời cảm ơn tới toàn bộ các thầy giáo, cô giáo của trường Đại học Công Nghệ - Đại học Quốc Gia Hà Nội nhưng người đã tạo điều kiện tốt nhất để tôi có thể học tập, nghiên cứu và hơn cả là đã truyền thụ cho tôi những hành trang kiến thức đầy đủ nhất.

Tôi cũng xin gửi lời cảm ơn chân thành nhất tới các thành viên trong nhóm nghiên cứu SkyLab, tới những người bạn người anh, chị đã giúp đỡ tôi hoàn thiện cả về kiến thức chuyên môn và kỹ năng học tập nghiên cứu.

Cuối cùng và không thể thiếu đó là lời cảm ơn tới bố mẹ và chị tôi và đặc biệt là bạn Dung Phùng những người đã luôn bên cạnh tôi giúp đỡ và động viên cổ vũ tinh thần tôi trong những lúc khó khăn nhất.

Tôi xin chân thành cảm ơn!

Hà Nội, ngày 24 tháng 04 năm 2019
Sinh viên

Nguyễn Tuấn Anh

TÓM TẮT

Khóa luận này trình bày một ứng dụng nhận dạng tên loài hoa dựa vào ảnh đầu vào trên nền tảng di động. Ứng dụng này được xây dựng với ba mô hình học máy chính: (1) một mô hình phân loại nhị phân để phát hiện có đối tượng hoa trong ảnh hay không, (2) một mô hình nhận dạng tên của loài hoa trong ảnh, và (3) một thuật toán tìm kiếm ảnh tương tự khi hiển thị các ảnh kết quả.

Ứng dụng được phát triển sử dụng phương pháp sử dụng lại một mô hình học sâu đã được huấn luyện trên bộ dữ liệu lớn khác để trích xuất dữ liệu của ảnh. Thực nghiệm so sánh phương pháp tiếp cận này với các cách trích xuất thuộc tính truyền thống chỉ ra tính khả thi và tiết kiệm của việc sử dụng lại mô hình giữa các bài toán nghiên cứu trong lĩnh vực xử lý ảnh.

Những đóng góp chính của khóa luận là: (1) Việt hoá tên cũng như đặc điểm về sinh trưởng, cách trồng của 102 loài hoa trong bộ dữ liệu Oxford-102, (2) phát triển ứng dụng trên nền tảng di động, và (3) phát triển mô hình phân loại ảnh có hoa hay không với độ chính xác 98.6

Từ khóa: nhận diện hoa, phát hiện hoa, ứng dụng di động

ABSTRACT

This thesis presents an recognition flowers' names application based on the input images on the mobile platform. This application is built with three main machine learning models: (1) a binary classification model to detect whether there is any flower in the image, (2) a model that recognizes the name of the flower in the image, and (3) an algorithm which searches for results of similar images.

The application is developed using a method of reusing a Deep learning models trained by another large data set to extract feature of image. Experimental comparison of this approach with traditional feature extraction method points to the feasibility and savings of reusing model of research problems in the field of image processing.

The main contributions of this thesis are: (1) Translating names, the growth characteristics as well as planting methods of 102 flower species in Oxford-102 data set, (2) Developing an application on mobile platform, and (3) Developing a flowers detection models with 98.6

Keywords: *flower detection, flower recognition, mobile application*

LỜI CAM ĐOAN

Tôi xin cam đoan toàn bộ khóa luận về ứng dụng phát triển phần mềm nhận dạng hoa trên thiết bị di động bao gồm mô hình nhận diện, phần mềm di động và phần mềm hệ thống là do tôi thực hiện dưới sự hướng dẫn của TS. Nguyễn Thị Ngọc Diệp và PGS. TS. Nguyễn Việt Hà.

Tất cả các công trình nghiên cứu, bài báo, khóa luận, tài liệu của các tác giả khác được tôi sử dụng trong khóa luận này đều được trích dẫn tường minh về tác giả và đều có trong danh sách tài liệu tham khảo.

Hà Nội, ngày 24 tháng 04 năm 2019

Sinh viên

Nguyễn Tuấn Anh

Mục lục

Danh sách bảng	vi
Danh sách ký hiệu, chữ viết tắt	vii
Danh sách hình vẽ	viii
Chương 1 Đặt vấn đề	1
Chương 2 Kiến thức cơ sở	2
2.1 Kỹ thuật thực thi tượng trưng	2
2.2 Bất biến vòng lặp	2
Chương 3 Công cụ VTSE kiểm chứng chương trình C/C++	3
3.1 Kiến trúc VTSE	3
Chương 4 Sinh bất biến vòng lặp	4
4.1 Kiến trúc InvaGen	4
Chương 5 Kết quả thực nghiệm	5
Chương 6 Kết luận	6

Danh sách bảng

Danh sách ký hiệu, chữ viết tắt

MRF	Markov Random Field
HSV	Hue, saturation, value
SIFT	scale-invariant feature transform
HOG	Histogram of Gradients
SVM	Support-vector machine
SVC	Support Vector Classification
LinearSVC	Linear Support Vector Classification
KNN	K-Nearest Neighbour
RGB	Red, green, blue
CNN	Convolutional Neural Network
CNNaug	Convolutional Neural Network + image augmentation
ONE	Online Nearest-neighbor Estimation
PCA	Principal Component Analysis

Danh sách hình vẽ

Chương 1

Đặt vấn đề

Các kĩ thuật CIA được thực hiện theo hai hướng tiếp cận chính là phân tích tĩnh (static CIA) và phân tích động (dynamic CIA) [2]. Trong thực tế, hầu hết các nghiên cứu đề xuất các phương pháp phân tích ảnh hưởng sự thay đổi liên quan đến static CIA. Nổi bật trong số đó là kĩ thuật CIA dựa trên sự phân loại thay đổi [4], kĩ thuật dựa vào đồ thị Call Graph [5] và kĩ thuật dựa trên tư tưởng giao thoa sóng nước WAVE-CIA [6]. Tuy nhiên, Các kĩ thuật CIA sử dụng đầu vào là kết quả của quá trình phân tích phụ thuộc từ ứng dụng. Quá trình này không thể tổng quát hóa cho toàn bộ các công nghệ và nền tảng hiện có.

Các phần còn lại của khóa luận được cấu trúc như sau. **Chương 2** trình bày về lý thuyết phân tích ảnh hưởng sự thay đổi. Cuối cùng, **Chương 6** là kết luận của toàn bộ khóa luận và công cụ, chương này cũng trình bày các công việc tiếp theo cần thực hiện.

Chương 2

Kiến thức cơ sở

Kỹ thuật thực thi tượng trưng

Bất biến vòng lặp

Chương 3

Công cụ VTSE kiểm chứng chương trình C/C++

Kiến trúc VTSE

Chương 4

Sinh bất biến vòng lặp

Kiến trúc InvaGen

Chương 5

Kết quả thực nghiệm

Chương 6

Kết luận

Tài liệu tham khảo

Tiếng Việt

- [1] Nguyễn Quỳnh Mai, Nguyễn Công Hưng. Phương pháp và công cụ phân tích sự ảnh hưởng của thay đổi mã nguồn cho các ứng dụng J2EE 6. *Hội nghị SVNCKH khoa CNTT, ĐH Công nghệ*, 2016.

Tiếng Anh

- [2] Bixin Li, Xiaobing Sun, Hareton Leung, Sai Zhang. A survey of code-based change impact analysis techniques. In *Software Testing, Verification and Reliability. Wiley Online Library* (2012).
- [3] Robert S. Arnold. 1996. Software Change Impact Analysis. IEEE Computer Society Press, Los Alamitos, CA, USA.
- [4] Xiaobing Sun, Bixin Li, Chuanqi Tao and Wanzhi Wen. Change Impact Analysis Based on a Taxonomy of Change Types. In *2010 IEEE 34th Annual Computer Software and Application Conference*, pages 373-382, 2010.
- [5] Linda Badri, Mourad Badri and Daniel St-Yves. Supporting Predictive Change Impact Analysis: A Control Call Graph Based Technique. In *Proceedings of the 12th Asia-Pacific Software Engineering Conference*, pages 167-175, 2005.

- [6] Bixin Li. WAVE-CIA: a novel CIA approach based on call graph mining. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 1000-1005, 2013.
- [7] Ba Cuong Le, Son Nguyen Van, Duc Anh Nguyen, Ngoc Hung Pham, Hieu Vo Dinh. JCIA: A Tool for Change Impact Analysis of Java EE Applications. *Information Systems Design and Intelligent Applications*, pp.105-114, 2018.
- [8] Don Brown, Chad Davis, and Scott Stanlick. 2008. Struts 2 in Action (In Action). Manning Publications Co., Greenwich, CT, USA.